REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
02/18/2012	Final	03/01/2009 - 11/30/2011

4. TITLE AND SUBTITLE
Dynamic Vision for Control

5a. CONTRACT NUMBER
FA9550-09-1-0427

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Stefano Soatto

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of California, Los Angeles

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR

AFOSR
875 N Randolph St
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-OSR-VA-TR-2012-0952

12. DISTRIBUTION/AVAILABILITY STATEMENT

DISTRIBUTION A: APPROVED FOR PUBLIC RELEASE

13. SUPPLEMENTARY NOTES

14. ABSTRACT

We have developed a comprehensive set of analytical and computational tools to ex- ploit visual data for the purpose of control and interaction with complex, dynamic and uncertain environments. The accomplishment of the goals set forth in the original proposal was articulated into three parallel research tracks. (1) Tracking; focused on the establishment of correspondence of low-level statistics across temporal samples, includ- ing the development of representations that are invariant to local illumination changes, co-variant with respect to finite-dimensional group transformations, and insensitive to non-invertible transformations due to non-group deformations, partial occlusions etc. (2) Motion Estimation: image motion established during tracking can be due to ego-motion, as well as to motion of independently moving objects in the scene. We have developed methods for multiple motion estimation and segmentation as well as techniques for integration of visual and inertial measurements that helped us exceed and push forward the state of the art in Visual SLAM (simultane- ous localization and mapping), which we have pioneered in years past.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	U	13	

DYNAMIC VISION FOR CONTROL
AFOSR FA9550-09-1-0427

Stefano Soatto
UCLA Vision Lab
University of California, Los Angeles

Final Report
February 18, 2012

Abstract

We have developed a comprehensive set of analytical and computational tools to exploit visual data for the purpose of control and interaction with complex, dynamic and uncertain environments. The accomplishment of the goals set forth in the original proposal was articulated into three parallel research tracks. (1) Tracking; focused on the establishment of correspondence of low-level statistics across temporal samples, including the development of representations that are invariant to local illumination changes, co-variant with respect to finite-dimensional group transformations, and insensitive to non-invertible transformations due to non-group deformations, partial occlusions etc. [28, 23, 18, 26, 30, 4, 17, 24, 1, 25]. (2) Motion Estimation: image motion established during tracking can be due to ego-motion, as well as to motion of independently moving objects in the scene. We have developed methods for multiple motion estimation and segmentation as well as techniques for integration of visual and inertial measurements that helped us exceed and push forward the state of the art in Visual SLAM (simultaneous localization and mapping), which we have pioneered in years past [2, 12, 16, 11, 29]. The two lines of work above have then been instrumental in (3) designing techniques for classifying and recognizing dynamic events from video [6, 15, 20, 21, 14]. The results of such a research program have been documented in a number of publications in the top journal and conference venues. In addition to targeted progress in the area above, during this project we have also developed basic image analysis tools for low-level processing [9, 19]. The software systems developed have been distributed worldwide through an open-source repository called VLFeat (www.vlfeat.org) that has become one of the standard libraries in industry, academia and government, together with the OpenCV.

1 Summary of Research Achievements

In this section we briefly summarize the technical achievements during this project. Details can be found in the published references.

1.1 Invariance in Representation

One of the central issues in processing visual data is to handle the large nuisance variability in the data: Images are affected by a large number of factors that are irrelevant for the task at hand. For instance, in decision tasks – say the detection, localization, recognition and categorization of an object or scene – vantage point is irrelevant, and so is illumination. In a control task – say tracking, docking, manipulation, etc. – reflectance properties of the scene are irrelevant, in addition of course to the more traditional nuisance factors such as spatial and range quantization, sensor noise, etc.

The Holy Grail would be the ability to infer from the data a "representation" that is at the same time invariant to nuisance factors, and "lossless" with respect to the task. In some cases, this is possible, as one can design statistics that are *invariant* to a particular nuisance and *sufficient* for a particular task. More often, however, one has to settle for a tradeoff between *insensitivity* to nuisances and *informativeness* for a particular task.

During the course of this project we have been able to precisely characterize the conditions under which it is possible to design a maximal invariant (to a nuisance) that is also a sufficient statistic (for a task). In [27] (with G. Sundaramoorthi, P. Petersen and V. S. Varadarajan), we have shown that even if the data (images) had infinite-resolution, and the nuisances (viewpoint and illumination) were drawn from an infinite-dimensional set, it is possible to extract an intermediate representation that (a) is invariant to the nuisance (so it contains *only* "information"), (b) is a sufficient statistic (so it contains *all* the "information"), and (c) it is discrete (it is supported on a zero-measure subset of the image domain). So, one can abstract discrete *"symbols"* from continuous data, and lose nothing when it comes to using it for decision and control.[1] In fact, the coding length of this internal representation is what I have suggested as a definition of some notion of information, called *Actionable Information*, following ideas that date back to J. J. Gibson [10], rather than the traditional notion of Information as Entropy of the data pioneered by Wiener and Shannon.

The resulting theory can be interpreted as a generalized *sampling theory* but not for the purpose of transmission and storage of data (as implicit in Shannon's theory), but for the purpose of *using* data for *decision and control tasks*. We have shown that under certain conditions one can take an infinite-dimensional signal (that is not band-limited, since there is no meaningful notion of band for sensing modalities subject to scaling phenomena) and reduce it to a finite set *without any loss of information*. This is, of course, not Shannon's information, as one would not be able to *reconstruct* the original signal. It is Gibson's information, in that the reduced representation is as good as the data for the purpose of any decision task that requires viewpoint and contrast invariance.

[1] Of course, if one were to use it for compression or transmission, the two tasks implicit in traditional Information Theory, then data analysis is by definition lossy.

While the construction in [27] works for any nuisance that has a *group structure*, for instance changes of viewpoint away from occlusions, and changes of illumination away from cast shadows, the latter *visibility artifacts* are not invertible, and therefore [27] cannot be applied. If object "A" is occluded by object "B" in an image, there is no processing of the image that will give us back object "A". A simple observation, that dates back to Gibson [10], resolve the conundrum are hold the key to enabling the development of a consistent theory of perception and action. Indeed, occlusion and quantizations are not invertible for a *passive observer*. However, *if one can control the sensing process*, then occlusions and quantization become invertible! Want to see object "A"? Just move around object "B". Want to resolve the fine structure in the far field? Move closer! This has enabled us to build on the theory of Actionable Information and establish a relation between the control authority of the sensing process and the gap between the Complete Information and the Actionable Information measured at the current time instant. Thus this "control-authority/actionable information" tradeoff extends "rate/distortion" theory when the underlying task is *not* the storage or transmission of data, but its *use* in decision and control tasks. This construction is described in [22].

This project has enabled us to establish a tight link between sensing and control, in the sense that *passive sensing* is subject to the usual limits imposed by traditional Information Theory. However, *active sensing* entailing control of the sensing process, enables closing the Actionable Information Gap. As Gibson put it in 1950, *we move in order to see, and we see in order to move.* The concept of Actionable Information is precisely the tie between sensing and control.

1.2 Occlusion detection and handling

There are two phenomena that affect that data formation process for imaging modalities that are critical in the analysis: *scaling* and *occlusion*. Scaling (due to changes of viewpoint under perspective imaging) causes the continuous limit to be part of the analysis (it is not possible to "discretize the world" and reduce the analysis to the discrete, because one can always move far enough away that any discretization is insufficient; conversely, the closer one get to an object or scene, the more details are being revealed, so the "source", to think in Communication terms, has infinite capacity). Occlusion is what makes *control* relevant. Consequently it should be no surprise that a significant portion of this research program has focused on occlusion handling and detection. The first breakthrough has been in the area of variational tracking.

Occlusion and clutter in variational tracking

In an influential 1989 paper, Mumford and Shah formalized the problem of segmenting an image (partitioning its domain into regions that exhibit smooth statistics) as a variational optimization problem. Their model has undergone numerous extensions and simplifications and is now widely used in applications ranging from tracking to medical imaging. The power of the Mumford-Shah model rests on the fact that it phrases a classification problem (clas-

sifying each point of the domain as either "target" or "background" where the target by definition occludes the background) as a regression problem (find the boundary by minimizing an energy functional). This can be formalized as a convex optimization, provided that there is one, and only one, target. Detection based on Mumford and Shah's approach finds a target even when none is present, and fails catastrophically when more than two regions are present. Several attempts to extend this approach to multiple regions or targets, the so-called "clutter problem", have been proposed, but have severe shortcomings. Some entail combinatorial optimization, others employ local searches based on heuristic choices of neighborhoods, and none preserves the convex nature of the optimization. In [29], we have drawn

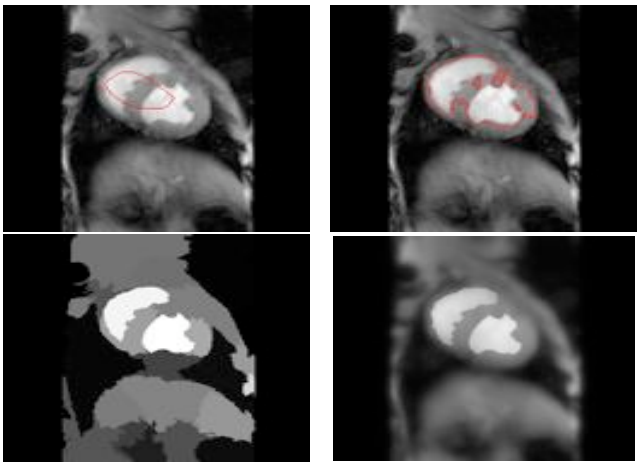

Figure 1: *Segmentation of a heart chamber using Chan & Vese's method (top-right, red curve), starting from the initial condition (top-left), is impeded by the fact that the background does not fit the constant model. Extension to multi-phase segmentation (bottom-left, each region is color-coded, and the object of interest corresponds to the white region) is complex and highly non-convex. Extension to more complex models, such as Mumford and Shah's (bottom-right) is also laborious. In both cases, precious modeling and computational resources are expended to capture the structure of the background away from the object of interest.*

from the literature on quickest set-point change detection to define a notion of locality that is controlled by the statistics of the data. This is illustrated in Fig. 2. Intuitively, considering a one-dimensional example (a "scan-line"), if the statistics in the region of interest are smooth, or at least continuous, then a discontinuity is well-defined and can be determined instantaneously (i.e., it is a point property). However, for a digital image that is everywhere discontinuous, discontinuity can be phrased as a hypothesis testing problem, and cannot be determined by considering an infinitesimal neighborhood. Instead, the smallest size of the neighborhood that can be considered for a given probability of error in the hypothesis test depends on the statistics of the "inside" region: The smoother the region, the smaller the "outlook" region that can be considered. This yields a model whereby the outlook region has an adaptive size that is regulated by the (estimated) statistics inside. This enables *decoupling*

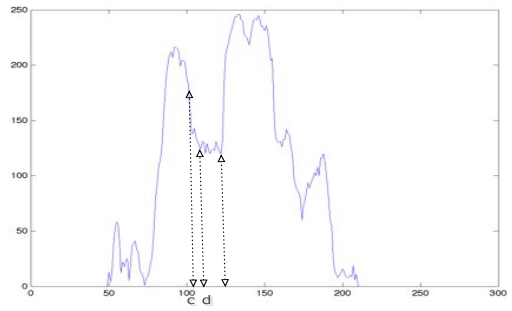

Figure 2: *One scanline from Fig. 1: The detection of the boundary c should be performed as soon as possible, d, so as not to have irrelevant background impinge on the decision (past the right-most dashed line).*

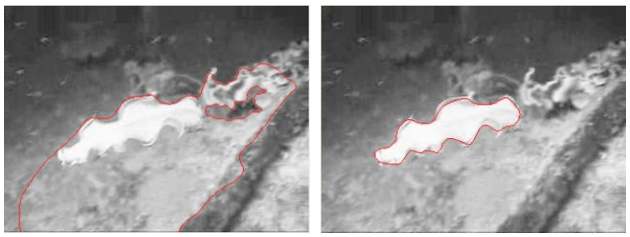

Figure 3: **Flatworm:** *The C-V model, as well as the full M-S model, fail to detect the boundaries of the flatworm. Our model, however, successfully detects it despite the complex background (right).*

multiple regions, and solving multiple convex problems on-line, where multiple initializations that converge to the same region are merged in a voting process. Unlike methods based on logical combinations of level set functions, local solutions do not affect each other, and can be computed in parallel. Figure 5 illustrates a representative example of comparison with classical active contours, and Figure 6 shows some quantitative comparisons.

While so far we have not specified what we mean by "statistics", and have implicitly assumed that the default statistic is the gray-scale level of the pixel, in natural images more complex statistics have to be considered, and they have to be defined at multiple scales. In [7] we have described distributional statistics and studied entropy regimes for multi-scale and stable analysis. In addition to static properties of the images, we have also commenced studying motion properties in this framework in [3].

Occlusion Detection

As already pointed out, *occlusion* phenomena play a central role in remote sensing, especially passive modalities such as EO and IR, where one has no control on the source signal (illumination). We have devoted considerable resources into the development of robust and efficient methods for *occlusion detection*, and we are pleased that our approach, presented in [1], has proven to exceed the state of the art both in terms of performance (precision/recall)

as well as computational efficiency.

In fact, we have shown that the problem of simultaneously estimating the indicator function of the occluded domain, as well as the domain deformation of the image (approximated by the optical flow under assumptions of Lambertian reflection and constant illumination) can be framed as a joint variational optimization problem that, under standard relaxation, can be shown to be a *convex* optimization problem. In [1] we have shown that the recently developed extended Lagrangian schemes known as "split Bregman methods" exceed the optimal (first-order) scheme due to Nesterov, both in terms of precision, as well as computational efficiency, by more than an order of magnitude. Our occlusion detection scheme has been distributed in source format and independently validated by other researchers.

Occlusion detection is important because it provide local cues of depth ordering, which in turn is critical for object detection, figure-ground segmentation, initialization of tracking etc. In particular, in [4] we have shown that, once occlusion detection between adjacent frames has been performed, global consistency cues can be integrated using *linear programming*! This yields an extremely efficient schemes for what we call *detachable object detection*, that is the detection of objects that are surrounded by the medium, except for their point of contact with the ground. This includes vehicles, people, animals, etc. This is the first time that such a difficult problem, that relates to motion segmentation, layer decomposition, and other notoriously difficult problems in dynamic visual processing, is shown to be solved using linear programming.

1.3 Filtering and prediction in the space of curves ("object-level filtering and prediction")

In order to integrate the results described above into a robust tracking framework, a model with predictive capabilities has to be employed. While this is standard in finite-dimensional state-spaces, deforming objects are best described as infinite-dimensional regions, or their boundaries. Therefore, a filter for an infinite-dimensional state space has to be designed. In [26] we have designed Luemberger-like observers for infinite-dimensional state-space that have a quotient structure under an infinite-dimensional Lie group (in the case of image domain deformations, this is the set of plane diffeomorphisms). Representative examples are shown in Figure 3.

Other contributions to this line of work include [20], where local templates are tracked independently and used for classification and time series, and [21], whereby the time series is assumed to be the output of a dynamical model, representing nuisance dynamics, and the "information" is encoded in the input, that is restricted to have sparse temporal gradients ("spikes"). Also along this line, in [8] we have proposed a filtering scheme to estimate, and then eliminate, the finite-dimensional group component in the data.

While a significant body of work has been devoted to tracking, some critical problems remained largely unsolved: The *initialization* problem, whereby one wishes to automatically detect multiple putative targets, without manual initialization [29], and the problem of predicting not only the coarse motion, but also the object-specific deformations [26]. We

have moved the state of the art forward by integrating occlusion detection into detachable object detection, and thence into tracking multiple deforming objects in the scene. This provides a complete description of all independently moving objects, organized in depth layers, from which the user can perform queries and/or select targets of interest.

Figure 4: *Detached object detection primes tracking in complex cluttered backgrounds.*

1.4 Vision-based navigation, mapping, localization

Our laboratory has pioneered the development of vision-based navigation, from the first ever demonstration of a real-time structure from motion system in 2000 (a system that takes live video from a regular camera and estimates three-dimensional trajectory of the camera as well as three-dimensional structure of the scene), to the latest visual-inertial integration system that has been recently published in the International Journal of Robotics Research. The system has been tested in open-loop on sequences up to more than 30Km with drift ranging from 0.1% to 0.5% of the traversed space. In addition, we have perfected the location recognition scheme that allows loop-closure and annihilation of the drift to within millimeter localization error, and the definition and real-time search of locations. The system has been implemented on an embedded platform and operates in real-time with up to tens of thousands of locations.

The final description of the system that we have been developing for the past 5 years is now complete and has appeared in print in [12]. We have also completed the software system CORVIS, and released an update (CORVIS2) that has been independently tested and validated.

1.5 Multiple Instance Filtering

In this latest development [30], we have developed an approach to filtering the state of a dynamical model that combines multiple-instance learning and semi-supervised learning. The basic premise is that modern tracking - unlike traditional tracking of point targets - can be framed as a learning problem, where one is given training sets (for instance, "examplars" or "samples" of what the target looks like, or simply a "bounding box" in the initial frame), and then wants to classify novel data for the presence, location, identity of (possibly multiple) targets. Unlike traditional tracking, the dynamics is not deterministic, but rather a prior in the detection problem.

Figure 5: ***Long Outdoor reconstruction.*** *Left: Our reconstruction of a 30 km long driving sequence, overlaid on an aerial view. Error is less than 0.5%. Right: Detail of area showing the position of point features and the motion reconstruction, overlaid to an orthographic aerial image.*

The challenge is that this problem does not fit the mold of traditional decision theory or machine learning, since the training set does not capture the variability under which the target is going to appear. One rarely has a training set of the target in all possible positions, orientations, poses, illumination, partial occlusion, etc. However, one would still like to detect the target under such a range of variability.

Tools from semi-supervised learning can be used to utilize all "labeled" data (e.g. given exemplar or bounding box) as well as all the given "unlabeled data" (images up to the "previous time" 't-1' where the target might be present, but its location is unknown) in order to classify the current frame (at time 't') by integrating them with assumptions on the dynamics of the target or the sensing platform.

In addition, the labeling can be imperfect: For instance, one often provides a "bounding box" of the target, that includes pixels-on-target as well as part of the background, and one rarely has a precise pixel-level segmentation of the target. Multiple-instance learning is a framework to exploit "weak labeling" where one is given negative samples (e.g. pixels that for sure are not on target, for instance those outside the bounding box) as well as a "positive bag" that contains some positive, but also some negative samples (e.g. the bounding box that contains pixels-on-target as well as pixels outside the target, without knowledge of which

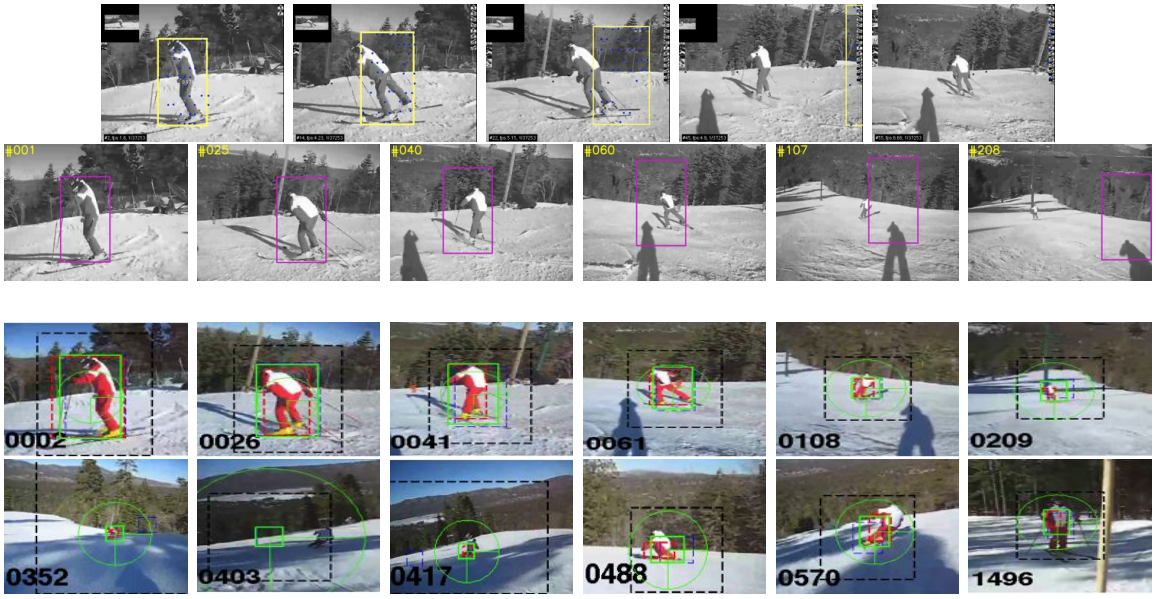

Figure 6: **Comparison of state-of-the-art trackers**: *The P/N Tracker [13] (first row)* **drifts** *because the target changes appearance and never returns to the initial configuration, and never recovers past frame 55. MIL Track [5] (second row)* **locks** *on a static portion of the background and fails at frame 208. Both phenomena are typical of tracking-by-detection approaches based on semi-supervised learning* without *explicit side-information. Our approach [30] maintains consisten track throughout the sequence despite large scale changes, changing background, and significant target deformation (third row). Of course, this approach fails too (***failure modes:** *bottom row), when the target is motion-blurred or subject to sudden illumination changes (frames 349 and 403 respectively) but quickly recovers (frames 352 and 417 respectively), missing 17 frames out of 1496 (98.86% tracking rate). For details see [30].*

is which).

We have integrated filtering in a classical non-linear point-estimate of the filtering density,with semi-supervised and multiple-instance learning, and shown that we can maintain tracking over long sequences for targets that are undergoing significant geometric, topological and photometric changes, despite a single "training set" consisting of a bounding box around the object of interest in the first frame. A representative set of results is shown in Fig. 6 in comparison with other state-of-the-art trackers.

Acknowledgment/Disclaimer

This work was sponsored in part by the Air Force Office of Scientific Research, USAF, under grant number FA9550-09-1-0427. The views and conclusions contained herein are those of the author and should not be interpreted as necessarily representing the official policies or

endorsements, either expressed or implied, of the Air Force Office of Scientific Research, or the U.S. Government.

References

[1] A. Ayvaci, M. Raptis, and S. Soatto. Optical flow and occlusion detection with convex optimization. In *Proc. of Neuro Information Processing Systems (NIPS)*, December 2010.

[2] A. Ayvaci, M. Raptis, and S. Soatto. Sparse occlusion detection with optical flow. *Intl. J. of Comp. Vision*, (in press) 2011.

[3] A. Ayvaci and S. Soatto. Motion segmentation with occlusions on the superpixel graph. In *Proc. of the Workshop on Dynamical Vision, Kyoto, Japan*, October 2009.

[4] A. Ayvaci and S. Soatto. Efficient model selection for detachable object detection. In *Proc. of EMMCVPR*, July 2011.

[5] B. Babenko, M.-H. Yang, and S. Belongie. Visual tracking with online multiple instance learning. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, 2009.

[6] A. Bissacco and S. Soatto. Hybrid dynamical models of human motion for the recognition of human gaits. *Intl. J. of Comp. Vis.*, 85(1):101–114, October 2009.

[7] S. Boltz, S. Soatto, and F. Nielsen. Entropy regimes for multi-scale and stable image analysis. In *Proc. of ECCV*, September 2010.

[8] A. Chiuso, G. Picci, and S. Soatto. *Communications in Information and Systems*, chapter Wide-sense Estimation on the Special Orthogonal Group. Springer, 2009.

[9] B. Fulkerson and S. Soatto. Really quick shift. In *Workshop on GPU For Computer Vision*, Crete, September 2010.

[10] J. Gibson. *The Perception of the Visual World*. Houghton Mifflin, 1950.

[11] B.-W. Hong and S. Soatto. Entropy-scale signatures for texture segmentation. In *Proc. of the Scale Space and Variational Methods Conference (SSVM)*, June 2011.

[12] E. Jones and S. Soatto. Visual-inertial navigation, localization and mapping: A scalable real-time large-scale approach. *Intl. J. of Robotics Research*, April 2011.

[13] Z. Kalal, K. Mikolajczyk, and J. Matas. Tracking-learning-detectino. *IEEE Trans. PAMI*, 6(1), 2011.

[14] T. Ko, D. Estrin, and S. Soatto. Warping background subtraction. In *Proc. of the IEEE Intl. Conf. on Comp. Vis. and Patt. Recog.*, 2010.

[15] T. Ko, S. Soatto, D. Estrin, and A. Cenedese. Cataloging birds in their natural habitat. In *Proc. of the ICPR Workshop on Vision in Natural Environments*, September 2010.

[16] T. Lee and S. Soatto. Edgel templates for fast object detection and pose estimation. In *Proc. of ISMAR*, May 27 2011.

[17] T. Lee and S. Soatto. Learning and matching multiscale template descriptors for real-time detection, localization and tracking. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, pages 1457–1464, 2011.

[18] T. Lee and S. Soatto. Video-based descriptors for object recognition. *Image and Vision Computing*, 2011.

[19] Y. Lou, P. Favaro, and S. Soatto. Nonlocal similarity image denoising. In *Proc. of the Intl. Conf. on Image Analysis and Processing (ICIAP)*, September 2009.

[20] M. Raptis and S. Soatto. Tracklet descriptors for action modeling and video analysis. In *Proc. of ECCV*, September 2010.

[21] M. Raptis, K. Wnuk, and S. Soatto. Spike train driven dynamical models for human motion. In *Proc. of the IEEE Intl. Conf. on Comp. Vis. and Patt. Recog.*, 2010.

[22] S. Soatto. Actionable information in vision. In *Proc. of the Intl. Conf. on Comp. Vision*, October 2009.

[23] S. Soatto. *Machine Learning for Computer Vision*, chapter Actionable Information in Vision. R. Cipolla, S. Battiato, G.-M. Farinella (Eds), Springer Verlag, 2011.

[24] S. Soatto and A. Chiuso. Controlled recognition bounds for scaling and occlusion channels. In *Proc. of the Data Compression Conference*, March 2011.

[25] G. Sundaramoorthi, A. Mennucci, A. Yezzi, and S. Soatto. Tracking deforming objects by filtering and prediction in the space of curves. December 2009.

[26] G. Sundaramoorthi, A. Mennucci, A. Yezzi, and S. Soatto. A new geometric metric in the space of curves with applications to tracking deforming objects by prediction and filtering. *SIAM J. on Imaging Science (in press)*, 2010.

[27] G. Sundaramoorthi, P. Petersen, and S. Soatto. On the set of images modulo viewpoint and contrast changes. *TR090005*, Submitted, JMIV 2009.

[28] G. Sundaramoorthi, P. Petersen, V. S. Varadarajan, and S. Soatto. On the set of images modulo viewpoint and contrast changes. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, June 2009.

[29] G. Sundaramoorthi, S. Soatto, and A. Yezzi. Curious snakes: A minimum-latency solution to the cluttered background problem in active contours. In *Proc. of the IEEE Intl. Conf. on Comp. Vis. and Patt. Recog.*, 2010.

[30] K. Wnuk and S. Soatto. Multiple instance filtering. In *Proc. of NIPS*, 2011.

Personnel Supported During Duration of Grant

Stefano Soatto (PI)
Michalis Raptis (Graduate Student) now at Disney Research, Pittsburgh
Zhao Yi (Graduate Student) now at Google
Kamil Wnuk (Graduate Student) now at WireWax, London, U.K.
Alper Ayvaci (Graduate Student)
Jingming Dong (Graduate Student)
Vasiliy Karasev (Graduate Student)

Publications (see section "References" above)

Honors & Awards Received

International Computer Vision Summer School, Invited Speaker, Scicli, July 2011; **US-Sino Summer School in Vision, Learning and Pattern Recognition**; **NIPS Tutorials**, Invited Lecturer; **International Conference on Computer Vision,** Program Co-Chair, Barcelona, Spain, Nov. 2011; **International Journal of Computer Vision**, Associate Editor; **Journal of Mathematical Imaging and Vision**, Associate Editor; **Foundations and Trends in Graphics and Vision**, Associate Editor; **International Workshop on Information Theory in Computer Vision**, Keynote Speaker, Barcelona, Nov. 2011; **Robotic Vision Workshop**, Keynote Speaker, Barcelona, Nov. 2011; **ICRA Workshop on Mobile Manipulation**, Shanghai, May 2011; **ECE Colloquium**, Boston University, October 2010; **ONR China Lake** Distinguished Lecture Series, July 2011. **Azriel Rosenfel Distinguished Lecturer**, University of Maryland, College Park, 2010;
Keynote speaker, Workshop on Dynamical Vision (WDV), Kyoto, Japan, 2009;
Distinguished Lecture Series, University of California, Irvine, 2009;
Program Co-Chair, Intl. Conf. on Computer Vision (ICCV), 2011, Barcelona, Spain;

AFRL Point of Contact

Dr. Fariba Fahroo, Program Manager, Computational Mathematics, AFOSR/NL, 875 North Randolph Street, Suite 325, Room 3112, Arlington, VA 22203, (703) 696-8429, Fax (703) 696-8450, DSN 426-8429, fariba.fahroo@afosr.af.mil
Transitions: Software developed during the project was integrated within the open-source repository "VlFeat" (www.vlfeat.org). Software for visual-inertial navigation has been distributed to numerous academic and government laboratories, including ONR China Lake.
New Discoveries: Most publications have been supplemented by open-source code and distributed freely for non-commercial purposes. No patented disclosures were submitted as part of the research conducted in this project.